

# Exponential inequalities for the distribution tails of canonical $U$ - and $V$ -statistics of $\rho$ -mixing observations

I. S. Borisov and N. V. Volodko

## Abstract

The Hoeffding-type-inequalities are obtained for the distribution tails of *canonical* (degenerate)  $U$ - and  $V$ -statistics of an arbitrary order based on samples from a stationary sequence of observations satisfying  $\rho$ -mixing.

*key words:* stationary sequence of random variables,  $\rho$ -mixing, multiple orthogonal series, canonical  $U$ - and  $V$ -statistics, Hoeffding's inequality.

## 1. Introduction

In the paper, we obtain some upper bounds for the distribution tails of  $U$ - and  $V$ -statistics with canonical bounded kernels, based on samples of stationary observations under  $\rho$ -mixing. The exponential inequalities obtained are a natural generalization of the classical Hoeffding's inequality for the distribution tail of a sum of independent identically distributed bounded random variables. The approach of the present paper is well known: It is based on the kernel representation of the statistics under consideration as a multiple orthogonal series (for detail, see [5, 11, 15]). The results obtained in the present paper improve the corresponding results in [6].

Introduce basic definitions and notions.

Let  $X_1, X_2, \dots$  be a stationary sequence of random variables taking values in an arbitrary measurable space  $\{\mathfrak{X}, \mathcal{A}\}$ , with a common distribution  $F$ . In addition to the stationary sequence introduced above, we need an auxiliary sequence  $\{X_i^*\}$  consisting of independent copies of  $X_1$ .

Denote by  $L_2(\mathfrak{X}^m, F^m)$  the space of measurable functions  $f(t_1, \dots, t_m)$  defined on the corresponding Cartesian power of the space  $\{\mathfrak{X}, \mathcal{A}\}$  with the corresponding product-measure and satisfying the condition

$$\mathbb{E}f^2(X_1^*, \dots, X_m^*) < \infty.$$

**Definition 1.** A function  $f(t_1, \dots, t_m) \in L_2(\mathfrak{X}^m, F^m)$  is called *canonical* (or *completely degenerate*) if

$$\mathbb{E}f(t_1, \dots, t_{k-1}, X_k, t_{k+1}, \dots, t_m) = 0 \quad (1)$$

for every  $k \leq m$  and all  $t_j \in \mathfrak{X}$ .

Define a Von Mises statistic (or  $V$ -statistic) by the formula

$$V_n \equiv V_n(f) := n^{-m/2} \sum_{1 \leq j_1, \dots, j_m \leq n} f(X_{j_1}, \dots, X_{j_m}). \quad (2)$$

In the sequel, we consider only the statistics where the function  $f(t_1, \dots, t_m)$  (the so-called *kernel* of the statistic) is canonical. In this case, the corresponding Von Mises statistic

is also called *canonical*. For independent  $\{X_i\}$ , such statistics are studied during last sixty years (see the reference and examples of such statistics in [11]). In addition to  $V$ -statistics, the so-called  $U$ -statistics were studied as well:

$$U_n \equiv U_n(f) := n^{-m/2} \sum_{1 \leq i_1 \neq \dots \neq i_m \leq n} f(X_{i_1}, \dots, X_{i_m}). \quad (3)$$

Notice also that any  $U$ -statistic is represented as a finite linear combination of canonical  $U$ -statistics of orders from 1 to  $m$ . This representation is called *Hoeffding's decomposition* (see [11]).

For independent observations  $\{X_i\}$ , we give below a brief review of results directly connected with the subject of the present paper. In this connection, we would like to mention the results in [3, Theorem 1], [2, Proposition 2.2], [1, Theorem 7, Corollary 3], and [7, Theorem 3.3].

One of the first papers where exponential inequalities for the distribution tails of  $U$ -statistics are obtained, is the article by W. Hoeffding [8] although he considered non-degenerated  $U$ -statistics only. In this case, the value  $(n-m)!/n!$  equivalent to  $n^{-m}$  as  $n \rightarrow \infty$ , is used as the normalizing factor instead of  $n^{-m/2}$ . The following statement was proved in [8]:

$$\mathbb{P}(U - \mathbb{E}U \geq t) \leq e^{-2kt^2/(b-a)^2}, \quad (4)$$

where

$$U = (n-m)!/n! \sum_{1 \leq i_1 \neq \dots \neq i_m \leq n} f(X_{i_1}, \dots, X_{i_m}),$$

$a \leq f(t_1, \dots, t_m) \leq b$  and  $k = [n/m]$ . In the case  $m = 1$ , inequality (4) is usually called *Hoeffding's inequality* for sums of independent identically distributed bounded random variables. Notice that, in this case, the sums mentioned may be simultaneously considered as canonical or nondegenerate  $U$ -statistics.

In [3], an improvement of (4) was obtained for the case when there exists a splitting majorant of the canonical kernel under consideration:

$$|f(t_1, \dots, t_m)| \leq \prod_{i \leq m} g(t_i), \quad (5)$$

and the function  $g(t)$  satisfies the condition

$$\mathbb{E}g(X_1)^k \leq \sigma^2 L^{k-2} k! / 2$$

for all  $k \geq 2$ . In this case, the following analogue of Bernstein's inequality holds:

$$\mathbb{P}(|V_n| \geq t) \leq c_1 \exp \left( - \frac{c_2 t^{2/m}}{\sigma^2 + L t^{1/m} n^{-1/2}} \right), \quad (6)$$

where the constants  $c_1$  and  $c_2$  depend only on  $m$ . Moreover, as noted in [3], inequality (6) cannot be improved in a sense.

It is clear that if  $\sup_{t_i} |f(t_1, \dots, t_m)| = B < \infty$  then, in (6), one can set  $\sigma = L = B^{1/m}$ . Then it suffices to consider only the deviation zone  $|t| \leq Bn^{m/2}$  in (6) (otherwise the left-hand side of (6) vanishes). Therefore, for all  $t \geq 0$ , inequality (6) yields the upper bound

$$\mathbb{P}(|V_n| \geq t) \leq c_1 \exp\left(-\frac{c_2}{2}(t/B)^{2/m}\right) \quad (7)$$

which is an analogue of Hoeffding's inequality (4).

In [2], an inequality close to (6) is proved without condition (5), and relation (7) is given as a consequence as well. In [7], some refinement of (7) is obtained for  $m = 2$ , and in [1], the later result was extended to canonical  $U$ -statistics of an arbitrary order.

The goal of the present paper is to extend inequality (7) to the case of stationary sequence of random variables  $\{X_i\}$  satisfying  $\rho$ -mixing. For dependent observations, we do not yet know how to get more precise inequalities close to Bernstein's inequality (6), for unbounded kernels under some moment restrictions only.

## 2. The main results for dependent observations

In the sequel, we assume that  $\mathfrak{X}$  is a separable metric space equipped with the Borel  $\sigma$ -field  $\mathcal{A}$ . Then the Hilbert space  $L_2(\mathfrak{X}, F)$  has a countable orthonormal basis  $\{e_i(t)\}$ . Put  $e_0(t) \equiv 1$ . Using the Gram–Schmidt orthogonalization, one can construct an orthonormal basis in  $L_2(\mathfrak{X}, F)$  containing the constant function  $e_0(t) \equiv 1$ . Then  $\mathbb{E}e_i(X_1) = 0$  for every  $i \geq 1$  due to orthogonality of all the other basis elements to the function  $e_0(t)$ . The normalizing condition means that  $\mathbb{E}e_i^2(X_1) = 1$  for all  $i \geq 1$ . It is well-known that the collection of functions

$$\{e_{i_1}(t_1)e_{i_2}(t_2)\cdots e_{i_m}(t_m); \quad i_1, \dots, i_m = 0, 1, \dots\}$$

is an orthonormal basis of the Hilbert space  $L_2(\mathfrak{X}^m, F^m)$ . The kernel  $f(t_1, \dots, t_m)$  can be decomposed by the basis  $\{e_{i_1}(t_1)\cdots e_{i_m}(t_m)\}$  and represented as the series

$$f(t_1, \dots, t_m) = \sum_{i_1, \dots, i_m=1}^{\infty} f_{i_1, \dots, i_m} e_{i_1}(t_1) \cdots e_{i_m}(t_m), \quad (8)$$

which converges in the norm of  $L_2(\mathfrak{X}^m, F^m)$  and the array  $\{f_{i_1, \dots, i_m}\}$  is square-summable at that. Note that the constant function  $e_0(t)$  is absent in representation (8) because the kernel is canonical (for detail, see [5]). Moreover, if the coefficients  $\{f_{i_1, \dots, i_m}\}$  are absolutely summable then, due to the B. Levi theorem and the simple estimate  $\mathbb{E}|e_{i_1}(X_1^*)\cdots e_{i_m}(X_m^*)| \leq 1$ , the series in (8) absolutely converges almost surely with respect to the distribution  $F^m$  of the vector  $(X_1^*, \dots, X_m^*)$ . In other words, if we substitute the random vector  $(X_1^*, \dots, X_m^*)$  for  $(t_1, \dots, t_m)$  in (8) then the equality in (8) is valid with probability 1 and the series absolutely converges with probability 1 as well.

It is worth noting that, even under the above-mentioned conditions on the coefficient, we cannot extend the series representation in (8) almost surely w.r.t. the distribution of the vector  $(X_1, \dots, X_m)$  with *dependent coordinates* (see [5], [6]). To extend the above representation to dependent observations we need some regularity conditions either on joint

distributions of the random variables  $(X_1, \dots, X_m)$  or on the kernel  $f(\cdot)$  of the statistics under consideration (see Theorems 1 and 2 below). In particular, under the regularity conditions in Theorem 1, equality (8) is satisfied *everywhere*. Then the corresponding Von Mises statistic (2) can be represented for all elementary events as follows:

$$\begin{aligned}
V_n &= n^{-m/2} \sum_{1 \leq j_1, \dots, j_m \leq n} f(X_{j_1}, \dots, X_{j_m}) \\
&= n^{-m/2} \sum_{1 \leq j_1, \dots, j_m \leq n} \sum_{i_1, \dots, i_m=1}^{\infty} f_{i_1, \dots, i_m} e_{i_1}(X_{j_1}) \cdots e_{i_m}(X_{j_m}) \\
&= \sum_{i_1, \dots, i_m=1}^{\infty} f_{i_1, \dots, i_m} n^{-1/2} \sum_{j=1}^n e_{i_1}(X_j) \cdots n^{-1/2} \sum_{j=1}^n e_{i_m}(X_j) \\
&= \sum_{i_1, \dots, i_m=1}^{\infty} f_{i_1, \dots, i_m} S_n(i_1) \cdots S_n(i_m),
\end{aligned}$$

where  $S_n(i_k) := n^{-1/2} \sum_{j=1}^n e_{i_k}(X_j)$ ,  $k = 1, \dots, m$ . Note that this representation is a base to prove limit theorems for such statistics.

In the present paper, we consider only stationary sequences  $\{X_j\}$  satisfying  $\rho$ -mixing condition. Recall the definition of this type of dependence. For  $j \leq k$ , denote by  $\mathfrak{M}_j^k$  the  $\sigma$ -field of all events generated by the random variables  $X_j, \dots, X_k$ , and by  $L_2(\mathfrak{M}_j^k)$  the space of all  $\mathfrak{M}_j^k$ -measurable random variables with finite second moments.

**Definition 2.** A sequence  $X_1, X_2, \dots$  satisfies  $\rho$ -mixing if

$$\rho(i) := \sup_{k \geq 1} \sup_{\xi \in L_2(\mathfrak{M}_1^k), \eta \in L_2(\mathfrak{M}_{k+i}^\infty)} \frac{|\mathbb{E} \xi \eta - \mathbb{E} \xi \mathbb{E} \eta|}{\mathbb{D} \xi \mathbb{D} \eta} \rightarrow 0, \quad i \rightarrow \infty,$$

where we set  $0/0 = 0$  by definition, and  $\mathbb{D}$  is the variance operator.

Recall that  $\rho$ -mixing coefficient is connected with the classical  $\alpha$ - and  $\varphi$ -mixing coefficients by the following two-sided inequality (see [9]):

$$4\alpha(i) \leq \rho(i) \leq 2\varphi^{1/2}(i).$$

If  $\{X_j\}$  is a stationary Gaussian sequence (or an arbitrary bijection of stationary Gaussian observations) then (see [10])

$$4\alpha(i) \leq \rho(i) \leq 2\pi\alpha(i).$$

Introduce some additional restrictions on the mixing coefficient and the kernels of the statistics under consideration:

- (A)  $\sum_{i_1, \dots, i_m=1}^{\infty} |f_{i_1, \dots, i_m}| < \infty$  and  $\sup_{i,t} |e_i(t)| \leq C < \infty$ .
- (B)  $\rho(k) \leq c_0 e^{-c_1 k}$ , where  $c_1 > 0$ ,  $c_0 \geq 1$ .

If  $\{X_j\}$  is an arbitrary  $m$ -dependent stationary sequence then condition (B) is fulfilled. If  $\{X_j\}$  is a stationary Markov chain satisfying  $\varphi$ - or  $\rho$ -mixing then condition (B) is always fulfilled (see [14]).

The main results of the present paper are contained in the following two theorems.

**Theorem 1.** *Let a canonical kernel  $f(t_1, \dots, t_m)$  be continuous (in every argument) everywhere on  $\mathfrak{X}^m$ . Moreover, if  $e_k(t)$  are continuous and both conditions (A) and (B) are satisfied then the following inequality holds:*

$$\mathbb{P}(|V_n| > x) \leq \exp \left\{ -C_1 x^{2/m} / B(f) \right\}, \quad (9)$$

where  $B(f) := \left( C^m \sum_{i_1, \dots, i_m=1}^{\infty} |f_{i_1, \dots, i_m}| \right)^{2/m}$  and  $C_1 > 0$  depends only on  $c_0$  and  $c_1$ .

**Theorem 2.** *Let the sequence  $X_1, X_2, \dots$  satisfy the following condition:*

**(AC)** *For all pairwise distinct subscripts  $j_1, \dots, j_m$ , the distribution of  $(X_{j_1}, \dots, X_{j_m})$  is absolutely continuous with respect to the distribution of  $(X_1^*, \dots, X_m^*)$ .*

Moreover, if both conditions (A) and (B) are fulfilled then

$$\mathbb{P}(|U_n| > x) \leq \exp \left\{ -C_1 x^{2/m} / B(f) \right\}, \quad (10)$$

where the constant  $C_1$  is the same as in Theorem 1.

**R e m a r k.** In [5] and [6], the reader can find some counterexamples which expose that continuity of the kernel in Theorem 1, condition (AC) in Theorem 2, and the requirement of absolutely summability of the coefficients in (A) cannot be omitted to derive the upper bounds for the distribution tails under consideration.

### 3. Proof of the theorems

*Proof of Theorem 1.* Without loss of generality, we assume that the separable metric space  $\mathfrak{X}$  coincides with the support of the distribution  $F$ . The last means that  $\mathfrak{X}$  does not contain open balls with  $F$ -measure zero. Since all the basis elements  $e_k(t)$  in (8) are continuous and uniformly bounded in  $t$  and  $k$ , due to Lebesgue's dominated convergence theorem, the series in (8) is continuous if the coefficients  $f_{i_1, \dots, i_m}$  are absolutely summable. It is not difficult to see that, in this case, the equality in (8) turns into the identity on the all variables  $t_1, \dots, t_m$  because the equality of two continuous functions on an everywhere dense set implies their coincidence everywhere. So, in this case, one can substitute *arbitrarily dependent* observations for the nonrandom variables  $t_1, \dots, t_m$  in identity (8). Therefore, for *all elementary events*, the above-mentioned representation holds:

$$V_n(f) = \sum_{i_1, \dots, i_m=1}^{\infty} f_{i_1, \dots, i_m} S_n(i_1) \cdots S_n(i_m), \quad (11)$$

where, as above,  $S_n(i_k) := n^{-1/2} \sum_{j=1}^n e_{i_k}(X_j)$ ,  $k = 1, \dots, m$ .

The method of deriving the exponential inequalities for the tail probabilities is based on the following version of Chebyshev's inequality:

$$\mathbb{P}(|V_n(f)| \geq x) \leq \inf_{N=1,2,\dots} \frac{\mathbb{E} V_n^{2N}}{x^{2N}}. \quad (12)$$

Notice that one can find this approach, for example, in [12], [3], and [7].

So, consider an arbitrary even moment of the above-introduced V-statistic using representation (11):

$$\mathbb{E}V_n^{2N} = \sum_{i_1, \dots, i_{2mN}=1}^{\infty} f_{i_1, \dots, i_m} \cdots f_{i_{2mN-m+1}, \dots, i_{2mN}} \mathbb{E}S_n(i_1) \cdots S_n(i_{2mN}). \quad (13)$$

Further, we have

$$\begin{aligned} |\mathbb{E}S_n(i_1) \cdots S_n(i_{2mN})| &\leq n^{-mN} \sum_{j_1, \dots, j_{2mN} \leq n} |\mathbb{E}e_{i_1}(X_{j_1}) \cdots e_{i_{2mN}}(X_{j_{2mN}})| \\ &= n^{-mN} \sum_{r=1}^{2mN} \sum_{k_1 < \dots < k_r \leq n} \sum_{s_j(i): i \leq r, j \leq 2mN} |\mathbb{E}\nu_{k_1} \cdots \nu_{k_r}|, \end{aligned} \quad (14)$$

where

$$\nu_{k_i} = e_{i_1}^{s_1(i)}(X_{k_i}) \cdots e_{i_{2mN}}^{s_{2mN}(i)}(X_{k_i}), \quad s_j(i) \geq 0, \quad \sum_{j=1}^{2mN} s_j(i) > 0,$$

and

$$\sum_{i=1}^r \sum_{j=1}^{2mN} s_j(i) = 2mN.$$

Notice that, for fixed  $r$ , the number of all collections of  $s_j(i)$  such that  $i$  runs from 1 to  $r$  and  $j$  runs from 1 to  $2mN$ , coincides with the number of all different arrangements of  $2mN$  indistinguishable elements in  $r$  cells when every cell must contain at least one element. It is well known that this value equals

$$C_{2mN-1}^{r-1} := \frac{(2mN-1)!}{(2mN-r)!(r-1)!}.$$

**Lemma 1.** *If the sequence  $\{X_i\}$  satisfies  $\rho$ -mixing and restriction (A) is valid then, for every collection  $\{i_1, \dots, i_{2mN}\}$ , the following inequality holds:*

$$|\mathbb{E}S_n(i_1) \cdots S_n(i_{2mN})| \leq (\tilde{c}C^2mN)^{mN},$$

where  $\tilde{c}$  depends only on the constants  $c_0$  and  $c_1$ .

*Proof.* We will estimate every summand of the external sum over  $r$  in (14) (taking the normalizing factor  $n^{-mN}$  into account). The approach is quite analogous to that in the proof of the corresponding assertion in [4, Lemma 4].

If  $r \leq mN$  then the number of all collections  $k_1 < \dots < k_r \leq n$  equals  $C_n^r$  and does not exceed  $n^r \leq n^{mN}$ . Hence,

$$n^{-mN} \sum_{k_1 < \dots < k_r \leq n} \sum_{\{s_j(i)\}} |\mathbb{E}\nu_{k_1} \cdots \nu_{k_r}| \leq C^{2mN} C_{2mN-1}^{r-1}.$$

Now, let  $r > mN$ . Fix an arbitrary collection of  $s_j(i)$  and consider the inner subsum in (14):

$$\sum_{k_{v_1} < \dots < k_{v_2} \leq n} |\mathbb{E} \nu_{k_{v_1}} \cdots \nu_{k_{v_2}}|, \quad (15)$$

where  $v_1$  and  $v_2$  are natural, with  $1 \leq v_1 < v_2 \leq r$ ; here  $v := v_2 - v_1 + 1$  is the multiplicity of the corresponding multiple subsum, and the blocks  $\nu_{k_i}$  are defined as before. Denote

$$\sum_{i=v_1}^{v_2} \sum_{j=1}^{2mN} s_j(i) = K(v_1, v_2).$$

Notice that, for  $v_1 \leq l < v_2$ , we have

$$K(v_1, l) + K(l+1, v_2) = K(v_1, v_2).$$

In the sequel, we will call  $\nu_{k_i}$  a *short block* if

$$\sum_{j=1}^{2mN} s_j(i) = 1,$$

i. e.,  $\nu_{k_i} = e_{i_j}(X_{k_i})$  for some  $1 \leq j \leq 2mN$ . Note that, in the case  $r > mN$ , there is at least one short block in the multiple subsum (15). We now prove the following statement: If the number of short blocks in the summands of subsum (15) is no less than  $d \in \{1, \dots, v\}$  then the following estimate is valid:

$$\sum_{k_{v_1} < \dots < k_{v_2} \leq n} |\mathbb{E} \nu_{k_{v_1}} \cdots \nu_{k_{v_2}}| \leq c_2^{d/2-1/4} d^{d/2-1} n^{v-d/2} C^{K(v_1, v_2)}, \quad (16)$$

where  $c_2$  may be chosen as

$$c_2 = \max \left\{ 16, 16 \left( \frac{c_0 e^{-c_1}}{1 - e^{-c_1}} \right)^4, \left( \frac{4c_0 e^{c_1}}{c_1} \right)^2 \right\}.$$

Prove (16) by induction on  $d$  for all fixed  $v_1$  and  $v_2$  such that  $v \leq r$ . First, let  $d = 1$ , i. e., the moments in (15) contain at least one short block. Denote it by  $\nu_{k_l}$ , where  $k_{v_1} \leq k_l \leq k_{v_2}$ . In addition, denote

$$\|\xi\|_t := (\mathbb{E} |\xi|^t)^{1/t}.$$

Then the following estimate holds:

$$\begin{aligned}
& \sum_{k_{v_1} < \dots < k_{v_2} \leq n} |\mathbb{E} \nu_{k_{v_1}} \dots \nu_{k_{v_2}}| \\
& \leq \sum_{k_{v_1} < \dots < k_{v_2} \leq n} \rho(k_{l+1} - k_l) \|\nu_{k_{v_1}} \dots \nu_{k_l}\|_2 \|\nu_{k_{l+1}} \dots \nu_{k_{v_2}}\|_2 \\
& \quad + \sum_{k_{v_1} < \dots < k_l \leq n} |\mathbb{E} \nu_{k_{v_1}} \dots \nu_{k_l}| \sum_{k_{l+1} < \dots < k_{v_2} \leq n} \mathbb{E} |\nu_{k_{l+1}} \dots \nu_{k_{v_2}}| \\
& \leq n^{v-1} C^{K(v_1, v_2)} \sum_{i=1}^{\infty} \rho(i) \\
& \quad + n^{v_2-l} C^{K(l+1, v_2)} \sum_{k_{v_1} < \dots < k_l \leq n} \rho(k_l - k_{l-1}) \|\nu_{k_{v_1}} \dots \nu_{k_{l-1}}\|_2 \|\nu_{k_l}\|_2 \\
& \leq n^{v-1} C^{K(v_1, v_2)} \sum_{i=1}^{\infty} \rho(i) + n^{v_2-l} C^{K(l+1, v_2)} n^{l-v_1} C^{K(v_1, l)} \sum_{i=1}^{\infty} \rho(i) \\
& = 2n^{v-1} C^{K(v_1, v_2)} \sum_{i=1}^{\infty} \rho(i) \leq 2n^{v-1/2} C^{K(v_1, v_2)} \frac{c_0 e^{-c_1}}{1 - e^{-c_1}}.
\end{aligned}$$

The induction base is proved.

Now, let the inequality

$$\sum_{k_{v_1} < \dots < k_{v_2} \leq n} |\mathbb{E} \nu_{k_{v_1}} \dots \nu_{k_{v_2}}| \leq c_2^{z/2-1/4} z^{z/2-1} n^{v-z/2} C^{K(v_1, v_2)}$$

hold for all the possible numbers  $z < d$  of the short blocks and all the multiplicities  $v$ , and the moments in (15) contain no less than  $d$  short blocks. Denote these blocks by  $\nu_{k_{j_1}}, \dots, \nu_{k_{j_d}}$ . Consider  $d-1$  pairs of neighbor blocks of the type  $\nu_{k_{j_s}}, \nu_{k_{j_s+1}}$ ,  $s \leq d-1$ . Denote the differences between the subscripts in these pairs by  $t_1, \dots, t_{d-1}$  respectively. Among the summands in (15), select  $d-1$  classes (in general, intersecting). We have

$$\sum_{k_{v_1} < \dots < k_{v_2}} |\mathbb{E} \nu_{k_{v_1}} \dots \nu_{k_{v_2}}| \leq R_1 + \dots + R_{d-1},$$

where the subsum  $R_s$  is taken over the set of subscripts

$$I_s := \{(k_{v_1}, \dots, k_{v_2}) : k_{v_1} < \dots < k_{v_2} \leq n, \quad t_s = \max t_i\}.$$

We estimate every subsum  $R_s$  as follows:

$$\begin{aligned}
R_s & \leq \sum_{I_s} \rho(k_{j_s+1} - k_{j_s}) \|\nu_{k_{v_1}} \dots \nu_{k_{j_s}}\|_2 \|\nu_{k_{j_s+1}} \dots \nu_{k_{v_2}}\|_2 \\
& \quad + \sum_{k_{v_1} < \dots < k_{j_s}} |\mathbb{E} \nu_{k_{v_1}} \dots \nu_{k_{j_s}}| \sum_{k_{j_s+1} < \dots < k_{v_2}} |\mathbb{E} \nu_{k_{j_s+1}} \dots \nu_{k_{v_2}}|. \tag{17}
\end{aligned}$$



Consider the first sum in the right-hand side of (17). We have

$$\begin{aligned}
& \sum_{I_s} \rho(k_{j_s+1} - k_{j_s}) \|\nu_{k_{v_1}} \cdots \nu_{k_{j_s}}\|_2 \|\nu_{k_{j_s+1}} \cdots \nu_{k_{v_2}}\|_2 \\
& \leq C^{K(v_1, v_2)} \sum_{I_s} \rho(t_s) \leq C^{K(v_1, v_2)} n^{v-(d-1)} \sum_{t_i: t_i \leq t_s} \rho(t_s) \\
& \leq C^{K(v_1, v_2)} n^{v-d+1} \sum_{k=1}^n \rho(k) k^{d-2} \leq C^{K(v_1, v_2)} n^{v-d/2} \sum_{k=1}^n \rho(k) k^{d/2-1} \\
& \leq C^{K(v_1, v_2)} n^{v-d/2} \sum_{k=1}^{\infty} \rho(k) k^{d/2-1} \\
& \leq C^{K(v_1, v_2)} n^{v-d/2} c_0 \int_0^{\infty} e^{-c_1 t} (t+1)^{d/2-1} dt \\
& = C^{K(v_1, v_2)} n^{v-d/2} c_0 e^{c_1} \int_0^{\infty} e^{-c_1(t+1)} (t+1)^{d/2-1} dt \\
& \leq \frac{c_0 e^{c_1}}{c_1^{d/2}} C^{K(v_1, v_2)} n^{v-d/2} \Gamma\left(\frac{d}{2}\right) \leq \frac{8c_0 e^{c_1}}{(2c_1)^{d/2}} C^{K(v_1, v_2)} n^{v-d/2} d^{d/2-2}.
\end{aligned}$$

The last inequality holds due to the evident fact that  $\Gamma(t) \leq 2t^{t-2}$ .

Now, consider the product of sums on the right-hand side of (17). Let the summands in the first of these sums contain  $d_1$  short blocks selected above. Correspondingly, in the summands of the second sum, there are  $d - d_1$  selected short blocks. By construction, we have  $1 \leq d_1 \leq d - 1$ , so, for both the sums, we can apply the induction assumption. Hence we have

$$\begin{aligned}
& \sum_{k_{v_1} < \cdots < k_{j_s}} |\mathbb{E} \nu_{k_{v_1}} \cdots \nu_{k_{j_s}}| \sum_{k_{j_s+1} < \cdots < k_{v_2}} |\mathbb{E} \nu_{k_{j_s+1}} \cdots \nu_{k_{v_2}}| \\
& \leq c_2^{d_1/2-1/4} d_1^{d_1/2-1} C^{K(v_1, j_s)} n^{j_s-v_1+1-d_1/2} \\
& \quad \times c_2^{(d-d_1)/2-1/4} (d-d_1)^{(d-d_1)/2-1} C^{K(j_s+1, v_2)} n^{v_2-j_s-(d-d_1)/2} \\
& \leq c_2^{d/2-1/4} d^{d/2-2} C^{K(v_1, v_2)} n^{v-d/2}
\end{aligned}$$

due to the evident inequality

$$c_2^{-1/4} (d-d_1)^{(d-d_1)/2-1} d_1^{d_1/2-1} \leq d^{d/2-2}.$$

Combining the estimates for the two sums on the right-hand side of (17), we obtain

$$R_s \leq d^{d/2-2} C^{K(v_1, v_2)} n^{v-d/2} \left( c_2^{d/2-1/2} + \frac{8c_0 e^{c_1}}{(2c_1)^{d/2}} \right) \leq c_2^{d/2-1/4} d^{d/2-2} C^{K(v_1, v_2)} n^{v-d/2}.$$

Summing all the upper bounds for  $R_s$ , we conclude that

$$\begin{aligned}
\sum_{k_{v_1} < \cdots < k_{v_2}} |\mathbb{E} \nu_{k_{v_1}} \cdots \nu_{k_{v_2}}| & \leq (d-1) c_2^{d/2-1/4} d^{d/2-2} C^{K(v_1, v_2)} n^{v-d/2} \\
& \leq c_2^{d/2-1/4} d^{d/2-1} C^{K(v_1, v_2)} n^{v-d/2}
\end{aligned}$$

what required to be proved.

Further, if  $r > mN$  then, in the summands in (14), there are no less than  $2(r - mN)$  short blocks. Thus, setting in (16)  $v_1 := 1$ ,  $v_2 := r$ ,  $d := 2(r - mN)$ , and  $v := r$ , we obtain the following estimate:

$$\begin{aligned} n^{-mN} \sum_{k_1 < \dots < k_r \leq n} \sum_{s_j(i)} |\mathbb{E} \nu_{k_1} \cdots \nu_{k_r}| &\leq C_{2mN-1}^{r-1} n^{-mN} c_2^{d/2} d^{d/2-1} n^{v-d/2} C^{K(v_1, v_2)} \\ &\leq 2^{2mN-1} c_2^{mN} (2mN)^{mN-1} C^{2mN} = 2^{3mN-2} c_2^{mN} C^{2mN} (mN)^{mN-1}. \end{aligned}$$

Summing over all  $r$  from 1 to  $2mN$  in (14), we conclude that

$$|\mathbb{E} S_n(i_1) \cdots S_n(i_{2mN})| \leq (8c_2)^{mN} C^{2mN} (mN)^{mN}.$$

The lemma is proved.  $\square$

By this lemma, we estimate the even moment of the Von Mises statistic in (13) as follows:

$$\mathbb{E} V_n^{2N} \leq \left( \sum_{i_1, \dots, i_m=1}^{\infty} |f_{i_1, \dots, i_m}| \right)^{2N} (\tilde{c} C^2 mN)^{mN}.$$

Therefore, applying Chebyshev's inequality (12), we obtain the upper bound

$$\mathbb{P}(|V_n| > x) \leq x^{-2N} \left( \sum_{i_1, \dots, i_m=1}^{\infty} |f_{i_1, \dots, i_m}| C^m \right)^{2N} (\tilde{c} mN)^{mN}$$

for any natural  $N$ . Set  $N = \alpha x^{2/m}$  for some  $\alpha > 0$ . (For simplicity, let  $N$  be natural.) Then

$$\begin{aligned} \mathbb{P}(|V_n| > x) &\leq x^{-2N} \left( \sum_{i_1, \dots, i_m=1}^{\infty} |f_{i_1, \dots, i_m}| C^m \right)^{2N} (\tilde{c} m \alpha)^{mN} x^{2N} \\ &= \exp \{ \alpha m \log(c_4 m \alpha) x^{2/m} \}, \end{aligned}$$

where

$$c_4 = \tilde{c} \left( \sum_{i_1, \dots, i_m=1}^{\infty} |f_{i_1, \dots, i_m}| C^m \right)^{2/m} = \tilde{c} B(f).$$

It is easy to verify that the multiplier  $\alpha m \log(c_4 m \alpha)$  reaches its minimum at the point  $\alpha = (c_4 m e)^{-1}$ , and this minimal value equals  $-(c_4 e)^{-1}$ . Then

$$\mathbb{P}(|V_n| > x) \leq \exp \{ -(\tilde{c} B(f))^{-1} x^{2/m} \}$$

what required to be proved.  $\square$

*Proof of Theorem 2.* As noted above, the series in (8) converges almost surely with respect to the distribution of the vector  $(X_1^*, \dots, X_m^*)$  if the coefficients  $f_{i_1, \dots, i_m}$  are absolutely summable. It is clear that condition (AC) allows to claim the same for the distribution of the vector  $(X_{j_1}, \dots, X_{j_m})$  for every *pairwise distinct* subscripts  $j_1, \dots, j_m$ .

Since, in the summation set in the definition of  $U$ -statistics, all the subscripts are pairwise distinct, we can substitute the series in (8) for the kernel in expression (3). So we obtain the following representation:

$$U_n = n^{-m/2} \sum_{i_1, \dots, i_m=1}^{\infty} f_{i_1, \dots, i_m} \sum_{1 \leq j_1 \neq \dots \neq j_m \leq n} e_{i_1}(X_{j_1}) \cdots e_{i_m}(X_{j_m}).$$

Under condition (A), the proof repeats the previous one almost literally. Estimating the even moment of the statistic in the same way, we can obtain an upper bound for the multiple sum of mixed moments of the basis elements which appear as a result of raising the  $U$ -statistic to the corresponding power. The difference between evaluations of the right-hand side of (13) and this expression is as follows: Instead of the mixed moment on the right-hand side of (13) (Lemma 1) we should estimate the expectation

$$\begin{aligned} n^{-Nm} \mathbb{E} \left| \sum_{1 \leq j_1 \neq \dots \neq j_m \leq n} e_{i_1}(X_{j_1}) \cdots e_{i_m}(X_{j_m}) \cdots \right. \\ \left. \times \sum_{1 \leq j_{2mN-m+1} \neq \dots \neq j_{2mN} \leq n} e_{i_{2mN-m+1}}(X_{j_{2mN-m+1}}) \cdots e_{i_{2mN}}(X_{j_{2mN}}) \right|. \end{aligned}$$

But the number of summands in the above multiple sum is less than that in the analogous multiple sum on the right-hand side of (14). Thus the multiple sum in (14) is an upper bound for the above-mentioned mixed moment. Hence the expectation of normalized multiple sum above does not exceed the upper bound in (14). From here the statement of Theorem 2 immediately follows with the same constants as those in Theorem 1.  $\square$

## References

- [1] ADAMCZAK, R. (2006). Moment inequalities for  $U$ -statistics. *Ann. Probab.* **34** 2288–2314. MR2294982.
- [2] ARCONES, M. A. and GINÉ, E. (1993). Limit theorems for  $U$ -processes. *Ann. Probab.* **21** 1494–1542. MR1235426.
- [3] BORISOV, I. S. (1991). Approximation of distributions of von Mises statistics with multidimensional kernels. *Siberian Math. J.* **32** 554–566. MR1142066.
- [4] BORISOV, I. S. and BYSTROV, A. A. (2006). Limit theorems for the canonical von Mises statistics with dependent data. *Siberian Math. J.* **47** 980–989. MR2302840.

- [5] BORISOV, I. S. and VOLODKO, N. V. (2008). Orthogonal series and limit theorems for canonical  $U$ - and  $V$ -statistics of stationary connected observations. *Siberian Adv. Math.* **18** 244–259. MR2437480.
- [6] BORISOV, I. S. and VOLODKO, N. V. (2009). Limit theorems and exponential inequalities for canonical  $U$ - and  $V$ -statistics of dependent trials. *High Dimensional Probability V*, 108–130. Inst. Math. Stat. Collect. 5, Luminy, France.
- [7] GINÉ, E., LATAŁA, R. and ZINN, J. (2000). Exponential and moment inequalities for  $U$ -statistics. *High Dimensional Probability II*, 13–38. Progr. Probab., 47, Birkhäuser, Boston. MR1857312.
- [8] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR144363.
- [9] IBRAGIMOV, I. A. and LINNIK, Yu. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff Publishing Company, Groningen, The Netherlands. MR322926.
- [10] IBRAGIMOV, I. A. and ROZANOV, Yu. A. (1978). *Gaussian Random Processes*. Applications of Mathematics, Springer-Verlag, New York-Berlin. MR543837.
- [11] KOROLYUK, V. S. and BOROVSKIKH, Yu. V. (1994). *Theory of U-Statistics*. Kluwer Academic Publ., Dordrecht. MR1472486.
- [12] MAJOR, P. (1981). *Multiple Wiener–Ito Integrals*. *Lecture Notes in Math.* **849**. Springer, New York. MR611334.
- [13] PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, New York. MR388499.
- [14] ROSENBLATT, M. (1971). *Markov Processes. Structure and Asymptotic Behavior*. Springer, Berlin. MR329037.
- [15] RUBIN, H. and VITALE, R. (1980). Asymptotic distribution of symmetric statistics. *Ann. of Statis.* **8** 165–170.